Deep Learning on Spatiotemporal Raster Data for Environmental Conservation

An Analysis of the Environmental Impact of Data Centers in Northern Virginia

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Emily Huber

December 1, 2024

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Prof. Pedro Augusto P. Francisco, Department of Engineering and Society

Prof. Brianna Morrison, Department of Computer Science

Prof. Mai Dahshan, Department of Data Science

Introduction

Climate change is one of our generation's most pressing issues. Acknowledging that climate change will affect nearly every aspect of life, the 2015 Paris Agreement set goals to limit global warming below 1.5 degrees Celsius by the end of this century (Nations, n.d.). To achieve global sustainability goals, high emitting industries must decrease greenhouse gas usage to reduce environmental footprints. Growing consumer dependence on artificial intelligence (AI) based services and applications have pushed companies to rely heavily on data centers needed to train AI models. Colloquially known as "Data Center Alley", Northern Virginia hosts the largest concentration of data centers in the world (*Clicking Clean Virginia The Dirty Energy Powering Data Center Alley*, n.d.). 70% of global internet traffic passes through Data Center Alley ("Virginia," 2023)**.** Northern Virginia has a large technology industry specializing in government contracting and resource-intensive server farms (Rosati et al., 2023). Tax exemptions, cheap energy from Dominion Energy, proximity to Washington, DC and relationships with the US federal government draws large companies and government agencies to place data centers in Northern Virginia ("Virginia," 2023)**.**

I will explore the environmental impact of data centers in Northern Virginia to find out how data centers impact climate change and local communities. As climate change worsens, we must find ways to reduce the environmental impact of the technology sector. Data centers have advanced the artificial intelligence landscape, however negatively contribute to environmental issues through high greenhouse gas emissions and water consumption.

The technical dimension of my Capstone project has to do with the application of artificial intelligence to environmental conservation through the use of geospatial data. I aim to explore how deep learning can extract spatial correlation in 3D spatiotemporal raster datasets.

Advocates of AI and data centers include technology companies that profit from AI and consumers who want access to AI-based services and applications. Advocates of data centers in Northern Virginia state that the economic benefits outweigh environmental impacts (Mullin, 2023). Critics of data centers may include local governments, communities, and environmental agencies. The technical dimension connects to the human dimension through the impact of data centers on local communities and environments. Large AI systems require intensive data center resources for model training. Data centers consume tons of water and energy for maintenance and cooling. Energy consumption puts pressure on local energy grids and potentially raises energy costs. The large consumption of water can stress dry ecosystems and cause environmental harm. Additionally, large-scale data centers can be loud and ugly.

Through my technical Capstone and STS research paper, I will apply AI to environmental conservation and discuss the environmental impact of data centers required to train AI systems. I will explore and comment on the potentially paradoxical nature of using AI for sustainability, given that in order to train AI models, large data centers must devour natural resources.

Technical Topic Proposal: Deep Learning on Spatiotemporal Raster Data

Geospatial data is critical to understanding the geographic world around us. Often sourced from satellite imagery or aerial photography, 3D geospatial data can help track changes in vegetation, land use, and climate patterns. Raster geospatial data represents geoscientific information through grids of regularly sized pixels. Each cell in a raster matrix contains a data value or categorical variable (*Raster Data*, n.d.). 3D geospatial data recorded over time is often called spatiotemporal data. While 3D geospatial raster data includes one raster image, spatiotemporal raster datasets hold a raster image for every unit of time. Analyzing 3D spatiotemporal data can reveal geospatial trends over time. Possible applications include natural resource management, air quality monitoring, or climate change modeling (*Spatiotemporal Analysis*, 2016).

When handling 3D spatiotemporal datasets, a common tradeoff is between spatial resolution and storage. Spatial resolution is defined by the dimension of each cell in the raster. A raster with high resolution represents a small geographic area in each cell. Low resolution rasters have cells that represent large geographic areas (*Raster Data Models*, n.d.). An expected challenge is the extensive computational requirements needed for analyzing spatiotemporal data. Depending on the chosen spatial resolution, raster data often outpaces the computational resources offered by typical computer processing units (CPUs). Another potential challenge with spatiotemporal datasets is the difficulty of extracting complex data relationships through traditional statistical methods. Specialized software and programming libraries exist to help with geospatial statistical analysis, however the sheer size and complexity of 3D spatiotemporal geographical systems makes analysis difficult (Hamdi et al., 2022). Through my research, I aim to figure out whether the use of deep learning on 3D spatiotemporal raster data can make the statistical analysis process more successful or computationally efficient.

Through my Capstone technical report, I aim to answer the question: How can 3D spatiotemporal raster data fit into an existing deep learning model to capture spatial correlations over time in geoscience systems? To begin, a dataset relating to the field of environmental management and conservation will be chosen. Extensive statistical analysis on the chosen 3D spatiotemporal raster dataset will be done to understand the correlations and relationships between features. Performing statistical analysis before involving artificial intelligence is important to understand the baseline data representation and relationships. Once inferences are found in the data, the goal is to improve the statistical analysis process with deep learning. A literature review will be conducted to understand the state of the art of 3D spatiotemporal raster data usage in deep learning. Then, I will work to integrate the chosen 3D spatiotemporal raster dataset into an existing deep learning model. The goal of using deep learning on spatiotemporal data is to capture spatial correlations between features in the dataset. Deep learning may be able to extract complex relationships otherwise difficult to understand with typical statistical analysis.

STS Topic Proposal: The Environmental Impacts of Data Centers in Northern VA

I will explore the environmental impact of data centers in Northern Virginia to discover how data centers impact climate change and local communities. This is important because as data centers continue to dominate the Northern Virginia landscape, their sociotechnical impact must be studied to understand and prevent environmental and community damage. Data centers have material impacts on environmental health and require lots of electricity that often comes from local power grids reliant on fossil fuels. Thousands of gallons of water per day are needed to cool servers (Ipsen, n.d.). When evaluating the environmental impacts of data centers, it is important to take a holistic approach, including both energy and water footprints into consideration.

Data centers are designed for information technology (IT) equipment, not people. Filled with rows of IT equipment racks containing servers, storage devices, and networking equipment, these buildings often have no windows and low air circulation. IT devices generate heat that must be extracted from the building to maintain operation. Air conditioning circulating within the building cools IT equipment (U.S. Environmental Protection Agency, 2007). The International Energy Association (IEA) estimates that 1-1.5% of global electricity use comes from data center energy demand (Chakraborty, 2024). Data centers can use over 40 times more energy than conventional office buildings. Energy is used for operation of the IT equipment, cooling of the IT equipment, and power delivery. Usually the people responsible for purchasing and operating IT equipment are not the same people responsible for paying utility bills and facilitating power & cooling. This split incentive means those who control the IT equipment energy use have little incentive to reduce power consumption (U.S. Environmental Protection Agency, 2007).

Growth in the data center sector has led to increased energy consumption from the power and cooling infrastructure that supports the servers. Increased energy use has led to higher energy costs for business and government, increased greenhouse gas emissions, higher strain on local power grids to meet increased demand, and increased capital costs for data center expansion and construction (U.S. Environmental Protection Agency, 2007).

Water footprint is the industry standard for quantifying water use, measuring the amount of freshwater consumed and polluted (Ristic et al., 2015). Water is used for cooling onsite as well as off-site electricity generation (Chakraborty, 2024). The amount of cooling needed to maintain IT equipment can be 5-10 times the cooling used in office or meeting spaces (Ristic et al., 2015).

Though many tech companies have stated corporate sustainability goals related to energy consumption, most of the renewable energy progress to the data center industry is not happening in Virginia. Some argue that this is because data centers have few options for buying renewable energy. Two thirds of Virginia's energy market is controlled by Dominion Energy, which actively lobbies for increased natural gas and coal usage (*Clicking Clean Virginia The Dirty Energy Powering Data Center Alley*, n.d.). In 2023, only 5% of Dominion Energy's energy came from renewable resources (Dominion Energy, 2023). The rise of "Data Center Alley" has proved incredibly profitable for Dominion Energy, which derives a large portion of its long-term business strategy from data center electricity sales. The dire need to transition away from fossil fuels and the sheer scale of data center electricity use in Northern Virginia puts global importance on the energy operations of Dominion Energy (*Clicking Clean Virginia The Dirty Energy Powering Data Center Alley*, n.d.).

The environmental impact of data centers in Northern Virginia will be analyzed with a literature review. Evidence on the environmental footprint, characterized by energy and water consumption, will be collected. The environmental impact of data centers outside of Northern Virginia will be interpreted to apply to Northern Virginia-specific data centers. Evidence specific to Northern Virginia data centers will also be collected. Northern Virginia power grid and energy use will be analyzed to determine how data center energy use fits into the larger energy economic status in Virginia. All of this evidence will be analyzed to provide a comprehensive overview of the environmental impacts of data centers in Northern Virginia.

Conclusion

Artificial intelligence can be applied to detect and solve complex environmental issues. However, the very existence of AI has negative environmental impacts, due to the high resource consumption of data centers required for model training. Through my technical Capstone and STS research paper, I aim to apply AI for environmental good, while also discussing the potentially paradoxical nature of using AI towards sustainability, given AI's greedy use of natural resources during training in data centers. With the rise of artificial intelligence, it is important to apply AI for societal good, such as fighting the most pressing environmental challenges. However, it is also crucial to understand AI's environmental impacts.

My technical Capstone will explore how deep learning can be applied to environmental conservation through the use of spatiotemporal raster data. It is expected that deep learning will be able to extract spatial relationships in the chosen dataset. My STS research paper will uncover how data centers required to train large AI systems environmentally impact the Northern Virginia region. The expected results of my STS research paper are an analysis of the various ways data centers impact the environment in Northern Virginia. I expect to uncover how energy and water consumption impact local energy grids and watersheds. Once completed, my technical Capstone project and STS research paper will contribute to our understanding of the complex interactions between artificial intelligence, data centers, and the environment.

References

Chakraborty, S. (2024). *Towards A Comprehensive Assessment of AI's Environmental Impact* (arXiv:2405.14004). arXiv. http://arxiv.org/abs/2405.14004

Clicking Clean Virginia The Dirty Energy Powering Data Center Alley. (n.d.). OurEnergyPolicy. Retrieved October 6, 2024, from https://www.ourenergypolicy.org/resources/clicking-clean-virginia-the-dirty-energy-powering-data-center-alley/

Dominion Energy. (2023). *2023 annual report and annual report on form 10-K*. https://s2.q4cdn.com/510812146/files/doc_downloads/2024/2024/03/20/20/Dominion-Energy-2023-Annual-Report-and-Annual-Report-on-Form-10-K.pdf

Hamdi, A., Shaban, K., Erradi, A., Mohamed, A., Rumi, S. K., & Salim, F. D. (2022). Spatiotemporal data mining: A survey on challenges and open problems. *Artificial Intelligence Review*, *55*(2), 1441–1488. https://doi.org/10.1007/s10462-021-09994-y

Ipsen, H. A. (n.d.). *Catching the Cloud and Pinning It Down: The Social and Environmental Impacts of Data Centers* [M.A., Syracuse University]. Retrieved October 14, 2024, from https://www.proquest.com/docview/2066820596/abstract/F94D5BCDD6A54884PQ/1

Mullin, J. (2023). Virginia's Data Centers and Economic Development: The state's fast-growing data center industry continues to build on its early advantages. *Econ Focus*, *28*(2), 14–17.

Nations, U. (n.d.). *The Paris Agreement*. United Nations; United Nations. Retrieved October 30, 2024, from https://www.un.org/en/climatechange/paris-agreement

Raster Data. (n.d.). Retrieved November 15, 2024, from https://docs.qgis.org/2.18/en/docs/gentle_gis_introduction/raster_data.html

Raster Data Models. (n.d.). Retrieved November 15, 2024, from https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-01-raster-data-models.html

Ristic, B., Madani, K., & Makuch, Z. (2015). The Water Footprint of Data Centers. *Sustainability*, *7*(8), 11260–11284. https://doi.org/10.3390/su70811260

Rosati, C., James, A., & Metcalf, K. (2023). Data plantation: Northern Virginia and the territorialization of digital civilization in "the Internet Capital of the World." *Online Media and Global Communication*, *2*(2), 199–227. https://doi.org/10.1515/omgc-2023-0017

Spatiotemporal Analysis. (2016, August 9). Columbia University Mailman School of Public Health. https://www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis

U.S. Environmental Protection Agency. (2007). *Report to Congress on server and data center energy efficiency: Public Law 109-431*. Washington, DC: U.S. Environmental Protection Agency. https://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

Virginia: The Data Center Capital of the World. (2023). *Site Selection*, *68*(2), 36–36.